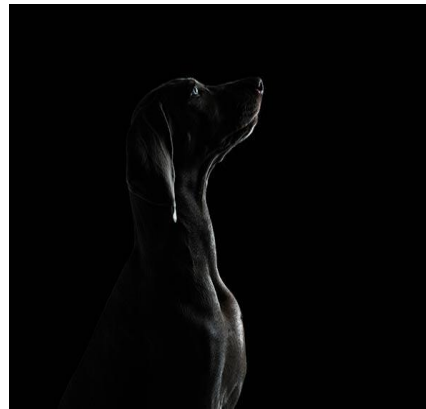
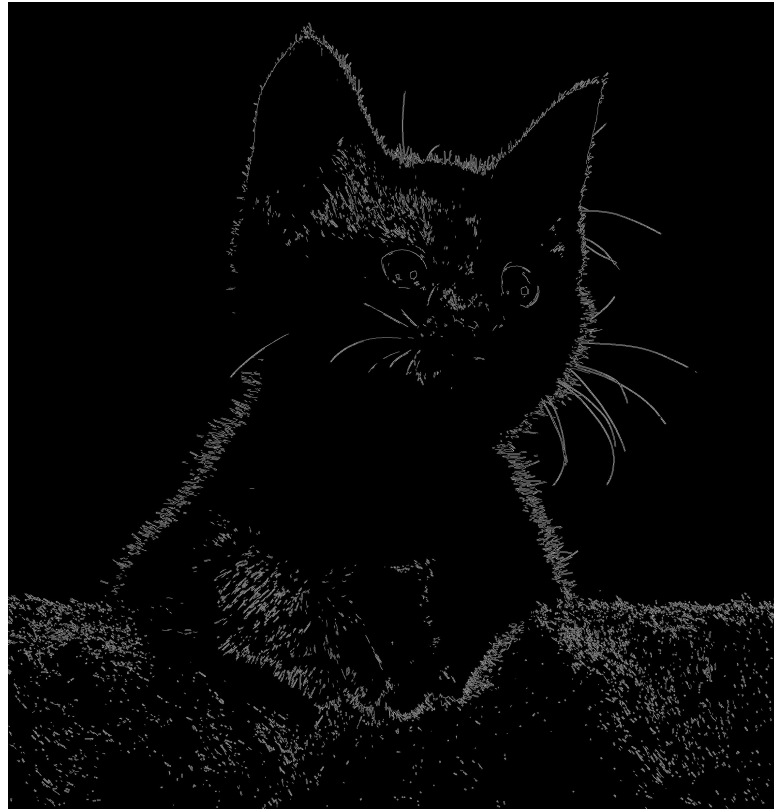




# Кошечки или собачки



# Алгоритм?

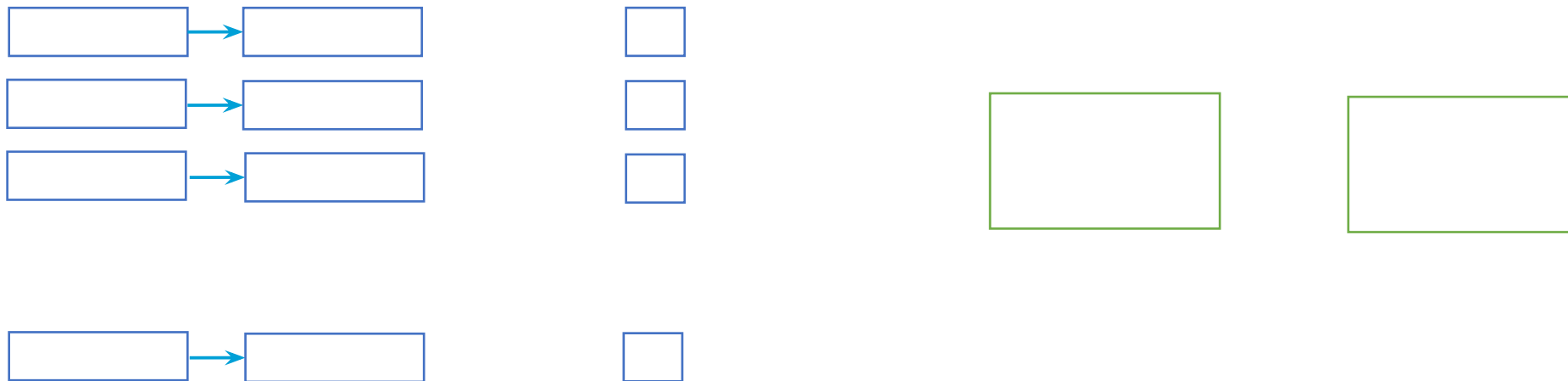


# Машинное обучение

# Machine Learning

**Machine learning** is a field of [computer science](#) that gives [computers](#) the ability to learn without being explicitly programmed.<sup>[1]</sup>

[Wikipedia](#)

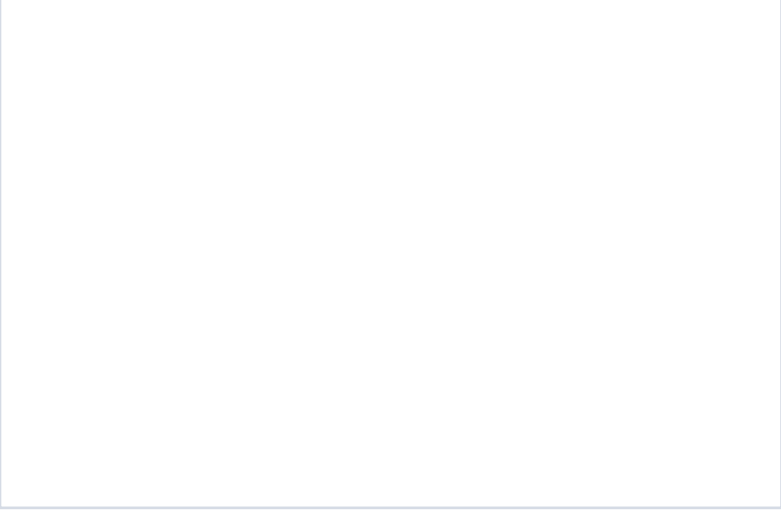




# Street View House Numbers



32x32  
pixels  
10 classes  
~70000  
train  
~25000  
test



# Метод ближайших соседей

## Nearest neighbor

1

20

Train:

просто все запомнить

Predict:

найти ближайший и  
выдать его класс

$$L_2 = \sqrt{\sum_i (v_i - u_i)^2}$$

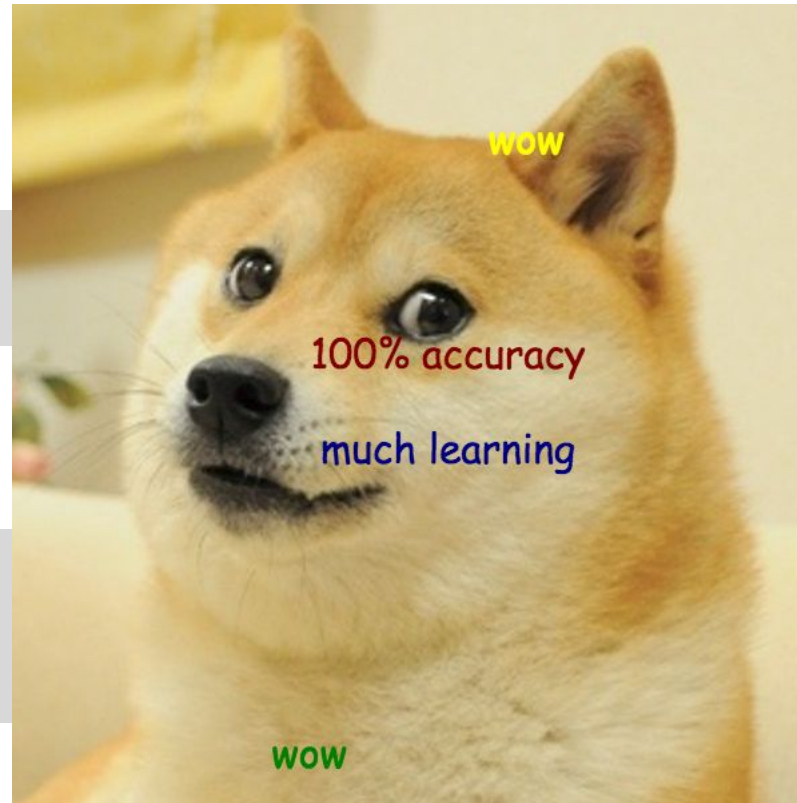
$$L_1 = \sum_i |v_i - u_i|$$

203

# Точность на тренировочных vs тестовых данных

Train:  
просто все запомнить

Predict:  
найти ближайший и  
выдать его класс



wow

20





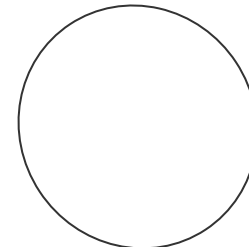
# Метод k-ближайших соседей

## K-nearest neighbors

1

20

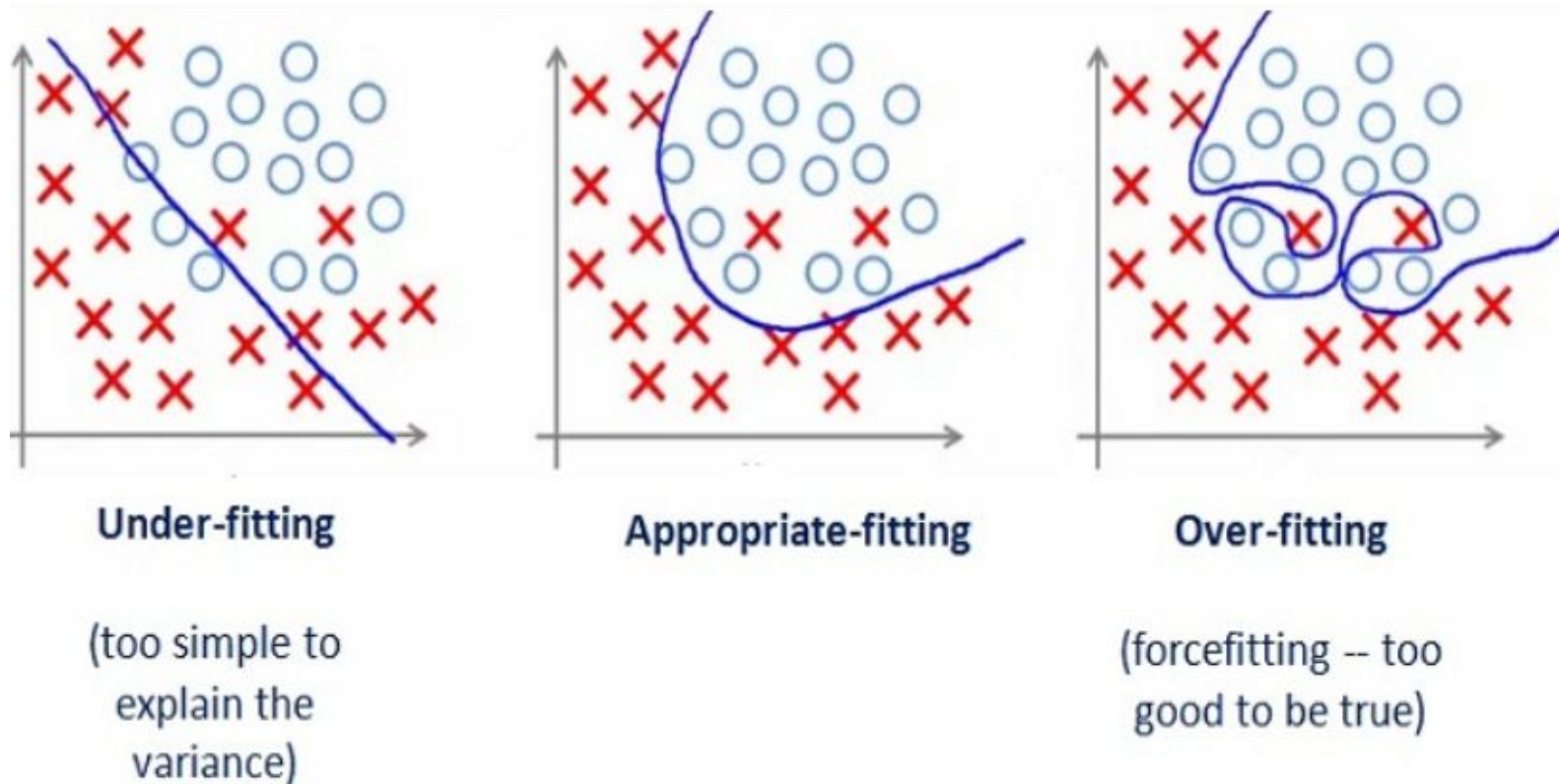
203



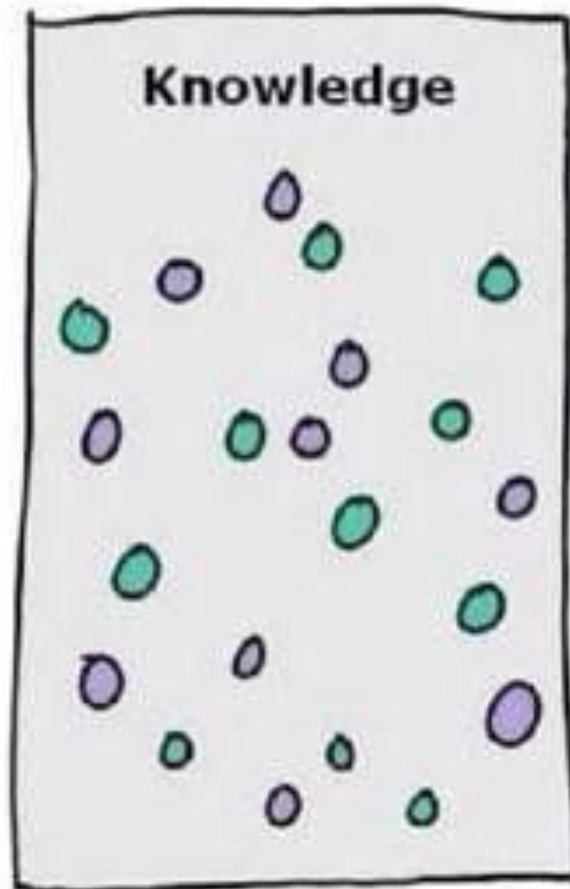
Как выбрать K?

# Переобучение и недообучение

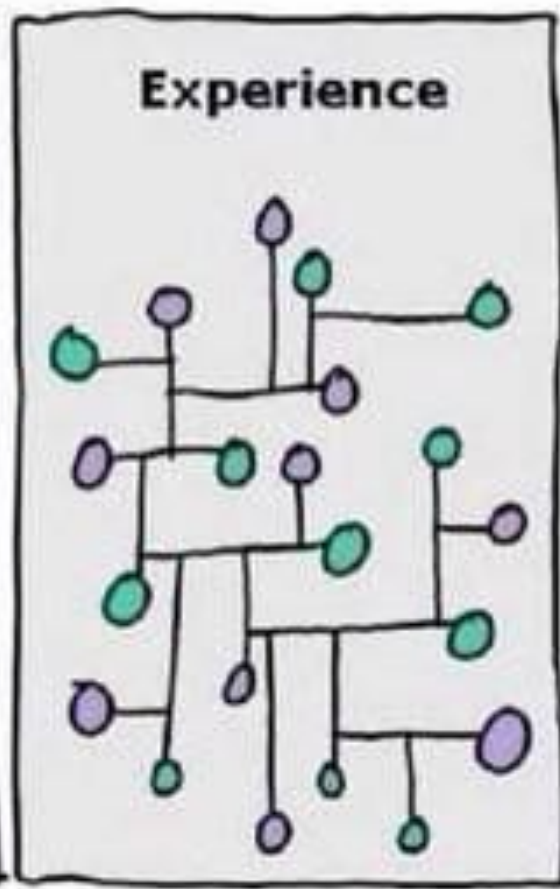
## Overfitting vs underfitting



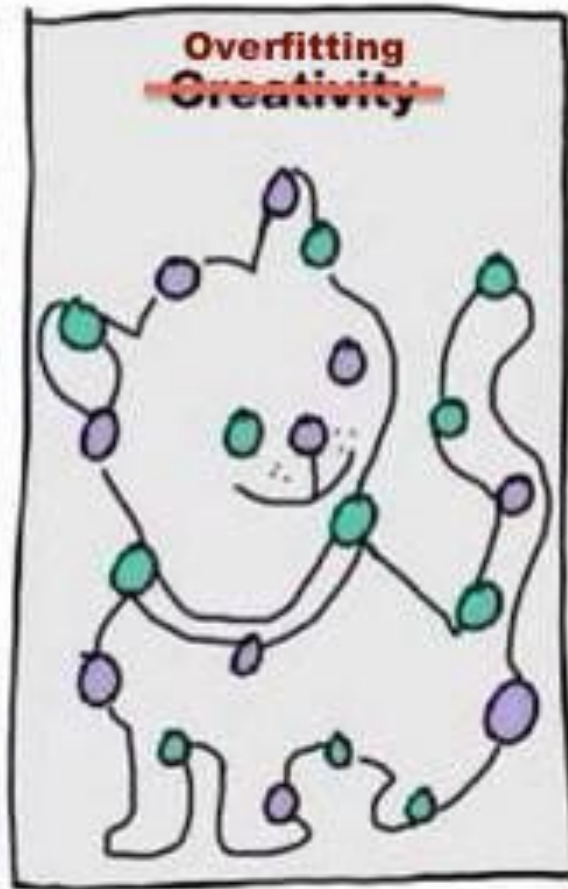
**Knowledge**

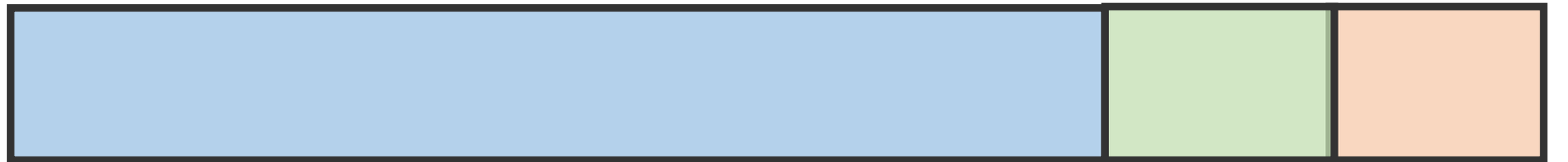
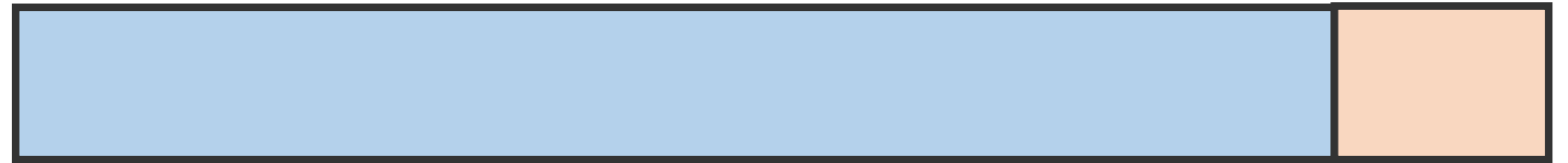
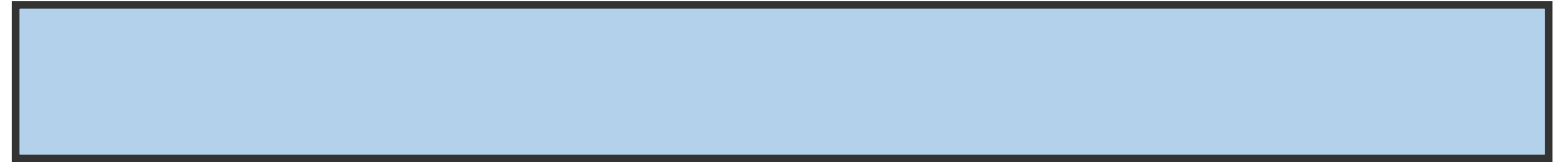
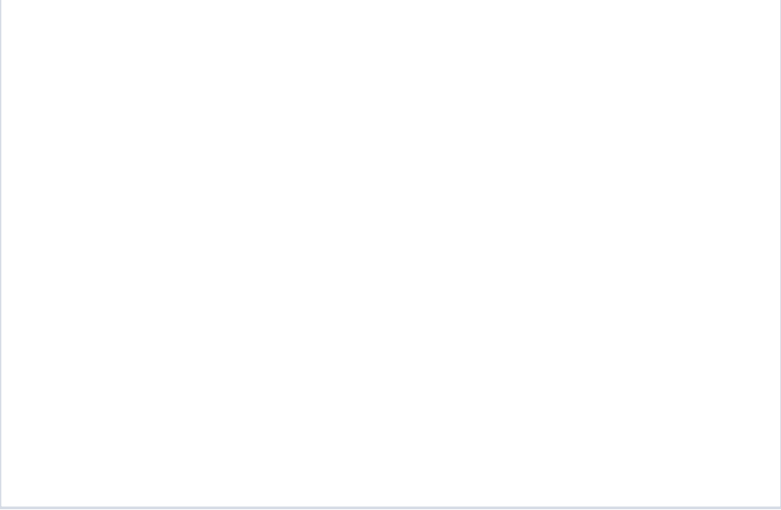


**Experience**



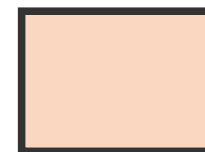
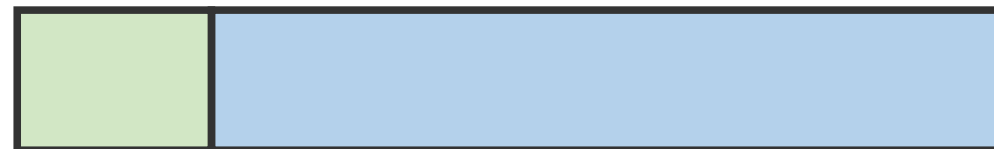
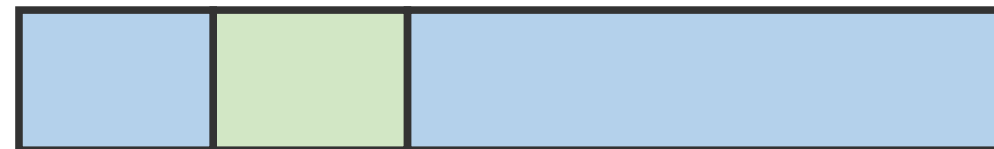
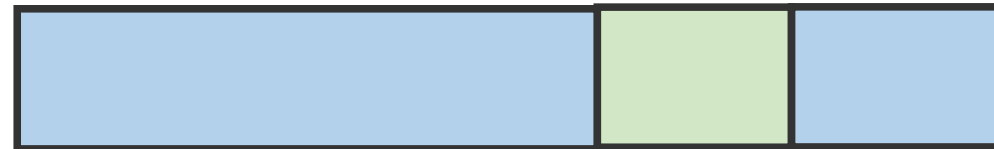
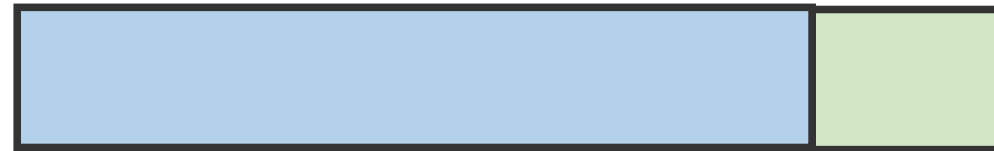
~~Overfitting~~  
~~Creativity~~







# Кросс-валидация Cross-validation



# Как сравнивать?

Бинарная классификация  
Binary classification



v  
s



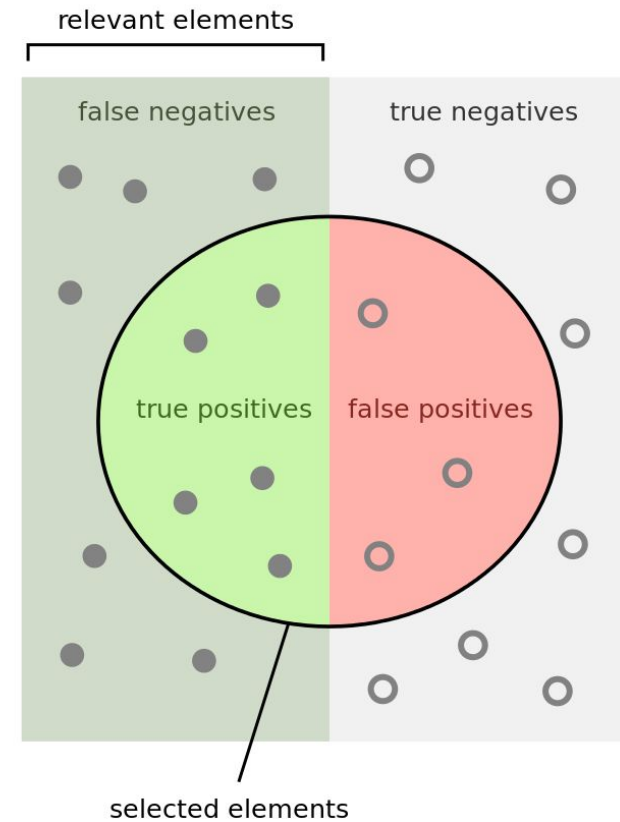
Точност  
ь

$$Accuracy = \frac{correct}{total}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} = \frac{2 * (precision * recall)}{precision + recall}$$



How many selected items are relevant?

$$Precision = \frac{\text{green semi-circle}}{\text{green and red semi-circles}}$$

How many relevant items are selected?

$$Recall = \frac{\text{green semi-circle}}{\text{green and light green rectangles}}$$

# Как сравнивать?

Бинарная классификация  
Binary classification



v  
s



Точность  
ь

$$Accuracy = \frac{correct}{total}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$



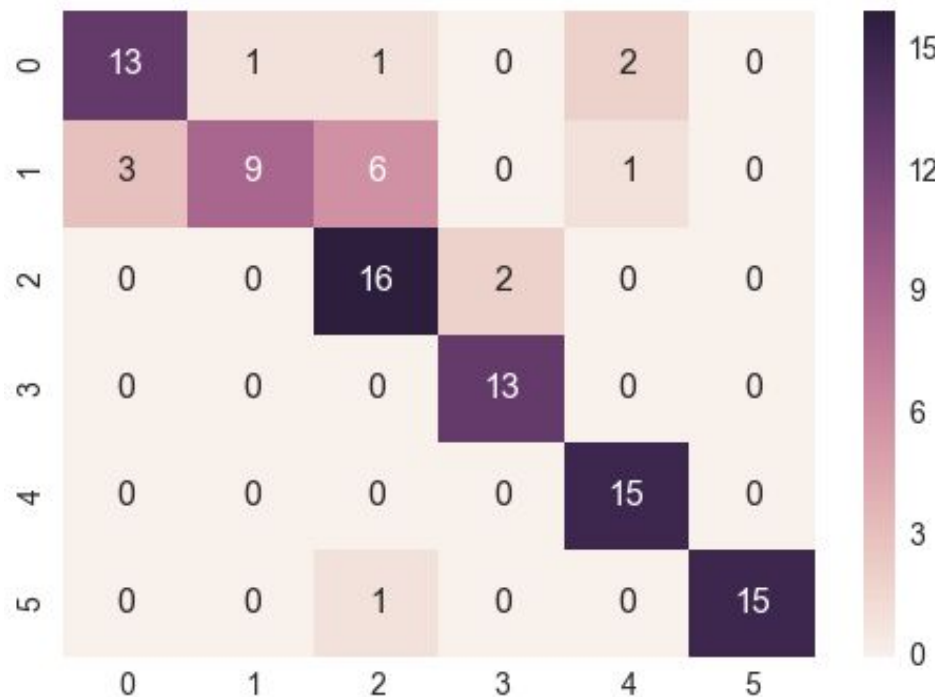
# Как сравнивать?

Многоклассовая классификация  
Multi-class classification



Точность  
ь

$$Accuracy = \frac{correct}{total}$$



$$Precision_c = \frac{A_{c,c}}{\sum_{i=1}^n A_{c,i}}$$

$$Recall_c = \frac{A_{c,c}}{\sum_{i=1}^n A_{i,c}}$$

$$Precision = \frac{\sum_{c=1}^n P_c}{n}$$

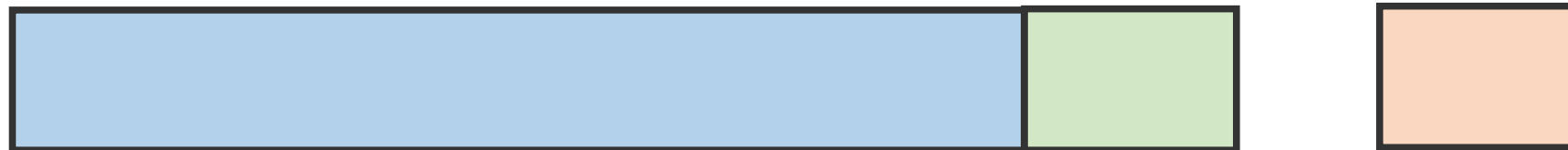
$$Recall = \frac{\sum_{c=1}^n R_c}{n}$$

[Stack](#)



MAKE GIFS AT [GIFSOUP.COM](http://GIFSOUP.COM)

# Machine Learning Flow



Ошибка на train

больша  
я

- Более мощную модель
- Больше ресурсов для тренировки
- Другой подход

маленька  
я

Ошибка на val

больша  
я

- Больше данных
- Больше регуляризации
- Другой подход

маленька  
я

Ошибка на test

больша  
я

- Отличаются train и test
- Больше данных, таких как test

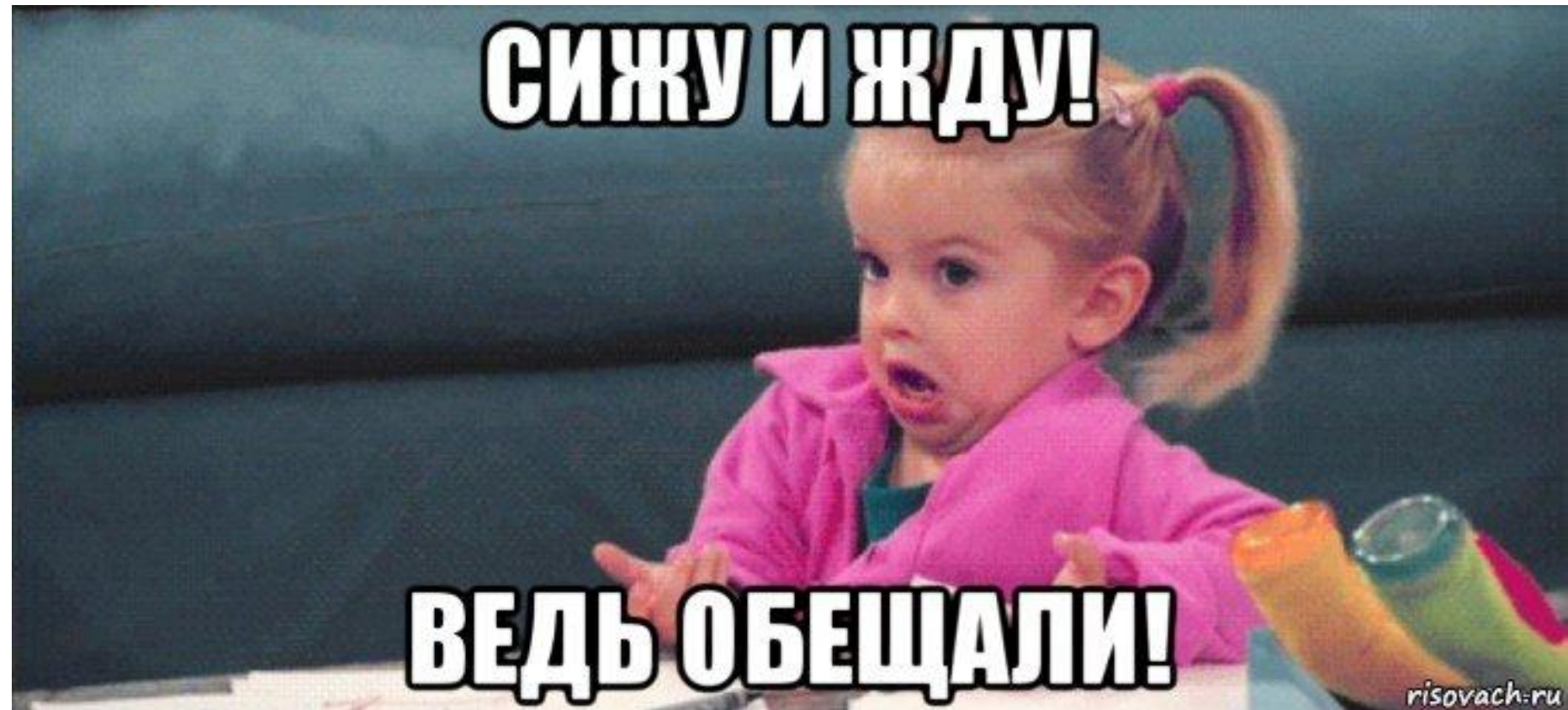
←  
маленька  
я

Кстати вот он про это подробнее:

[Nuts and Bolts of Applying Deep Learning](#)



В следующий раз уже будет  
про нейросети



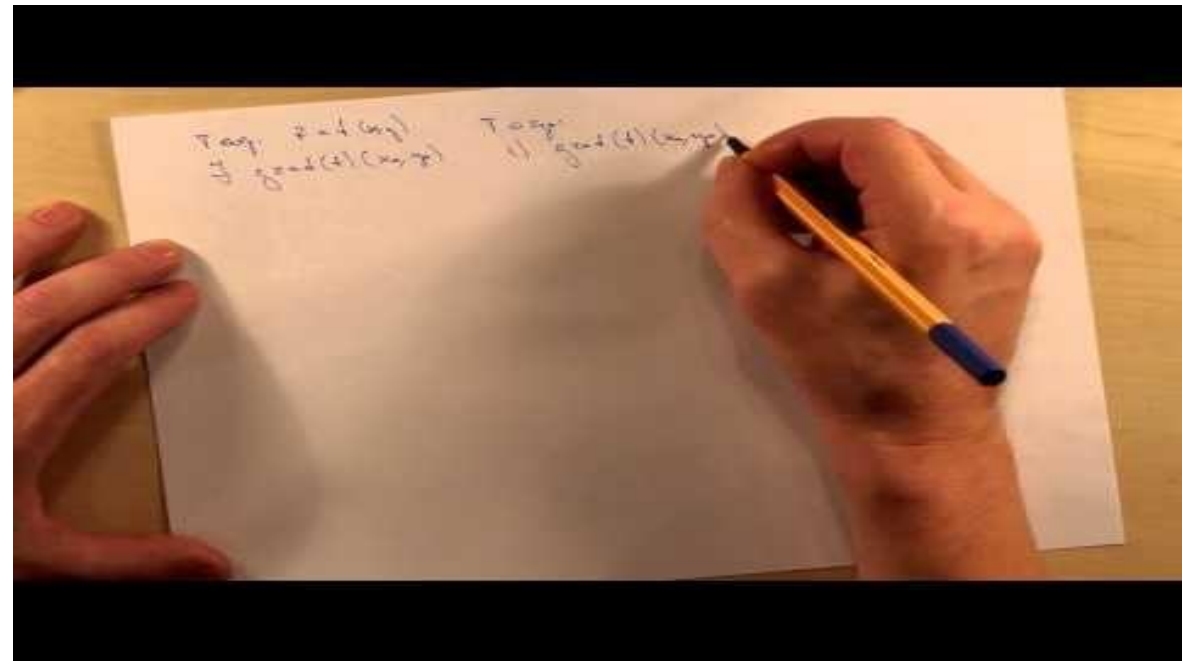


# Домашнее задание!

Повторить производную сложной функции (chain rule)



[Link](#)



[Link](#)